

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号
特開2001-34587
(P2001-34587A)

(43)公開日 平成13年2月9日(2001.2.9)

(51)Int.Cl. ⁷	識別記号	F I	テームコード*(参考)
G 0 6 F 15/167		G 0 6 F 15/167	B 5 B 0 4 5
			C
15/177	6 7 2	15/177	6 7 2 K

審査請求 有 請求項の数11 O L (全 11 頁)

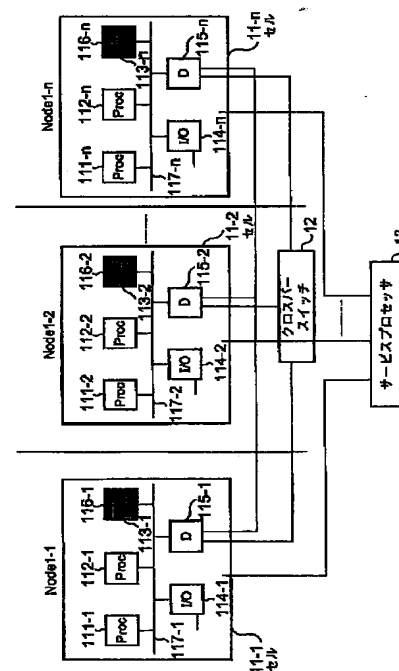
(21)出願番号	特願平11-203754	(71)出願人	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(22)出願日	平成11年7月16日(1999.7.16)	(72)発明者	門井 忠茂 東京都港区芝五丁目7番1号 日本電気株式会社内
		(74)代理人	100088328 弁理士 金田 暢之 (外2名) Fターム(参考) 5B045 BB44 DD01 JJ07 JJ48

(54)【発明の名称】 マルチプロセッサシステム

(57)【要約】

【課題】 メモリへのアクセスがノード固有メモリ空間及びノード間通信用メモリ空間のいずれへのアクセスであるかを容易に判別するとともに、メモリへのアクセス時に発生した障害がノード固有メモリ空間及びノード間通信用メモリ空間のいずれで発生した障害であるかを容易に、かつ迅速に判別する。

【解決手段】 メモリコントローラ116-1において、メモリ113-1へのアクセス時に、通信ドライバ115-1にて発行されたトランザクションに通信空間アクセスフラグがセットされていることを確認した場合にノード間通信用メモリ空間へのアクセスであると判断されるとともに、トランザクションが保持され、メモリ113-1へのアクセス時に障害が発生した時に、保持されたトランザクションに通信空間アクセスフラグがセットされていることを確認した場合に、ノード間通信用メモリ空間に障害が発生したと判別される。



【特許請求の範囲】

【請求項1】 複数のプロセッサと、該複数のプロセッサのそれぞれによってデータが記憶されるメモリとを具備するセルからなる複数のノードと、該複数のノードのいずれかにて前記メモリに障害が発生した場合に該複数のノードのそれぞれに対して障害時に処理内容を知するサービスプロセッサとを有し、前記複数のノードのそれぞれが、前記メモリにおけるメモリ空間の一部をノード間通信用メモリ空間と設定し、前記メモリにおける前記ノード間通信用メモリ空間にアクセスして互いにノード間通信を行うマルチプロセッサシステムにおいて、前記複数のノードのそれぞれは、前記ノード間通信用メモリ空間へのアクセス時に、通信空間アクセスフラグをセットしたトランザクションを発行する通信ドライバと、前記メモリへのアクセス時に、前記通信空間アクセスフラグを検出した場合に、前記メモリへのアクセスが前記ノード間通信用メモリ空間へのアクセスであると判断するメモリアクセス制御部とを有することを特徴とするマルチプロセッサシステム。

【請求項2】 請求項1に記載のマルチプロセッサシステムにおいて、前記メモリアクセス制御部は、前記通信ドライバにて発行されたトランザクションを保持するとともに、前記メモリへのアクセス時に発生した障害を検出し、保持された前記トランザクションに前記通信空間アクセスフラグがセットされていることを検出した場合に、前記メモリへのアクセス時に発生した障害が前記ノード間通信用メモリ空間における障害であると判断することを特徴とするマルチプロセッサシステム。

【請求項3】 請求項2に記載のマルチプロセッサシステムにおいて、前記メモリアクセス制御部は、前記ノード間通信用メモリ空間において障害が発生したと判断した場合に、前記サービスプロセッサが前記複数のノードのそれぞれに対して障害時の処理内容を知するための割り込み信号を生成して前記サービスプロセッサに対して出力し、前記サービスプロセッサは、前記割り込み信号が入力されると、該割り込み信号により障害の情報を採取し、該障害の情報に基づく処理内容を、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して通知することを特徴とするマルチプロセッサシステム。

【請求項4】 請求項3に記載のマルチプロセッサシステムにおいて、前記サービスプロセッサは、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して、ブロードキャスト割り込みを発生して前記処理内容を通知することを特徴とするマルチプロセッサシステム。

【請求項5】 請求項1乃至4のいずれか1項に記載のマルチプロセッサシステムにおいて、

前記複数のノードのそれぞれは、前記セルを複数個有し、該複数のセルのうち、いずれかのセルを分離させることを特徴とするマルチプロセッサシステム。

【請求項6】 請求項3乃至5のいずれか1項に記載のマルチプロセッサシステムにおいて、

前記メモリアクセス制御部は、前記通信ドライバにて発行されたトランザクションを保持するトランザクション保持回路と、前記メモリへのアクセス時に発生する障害の検出を行うメモリバリティチェック回路と、

前記トランザクション保持回路に保持されたトランザクションに前記通信空間アクセスフラグがセットされているか検出するとともに、前記メモリバリティチェック回路にて障害が検出された場合に、前記メモリに発生した障害が前記ノード間通信用メモリ空間における障害であると判断する障害判断部と、

前記障害判断部において前記ノード間通信用メモリ空間に障害が発生したと判断された場合に、前記割り込み信号を生成して前記サービスプロセッサに対して出力するメモリ障害割り込みラッチ回路とを有することを特徴とするマルチプロセッサシステム。

【請求項7】 複数のプロセッサと、該複数のプロセッサのそれぞれによってデータが記憶されるメモリとを具備するセルからなる複数のノードと、該複数のノードのいずれかにて前記メモリに障害が発生した場合に該複数のノードのそれぞれに対して障害時に処理内容を知するサービスプロセッサとを有し、前記複数のノードのそれぞれが、前記メモリにおけるメモリ空間の一部をノード間通信用メモリ空間と設定し、該ノード間通信用メモリ空間にアクセスして互いにノード間通信を行うマルチプロセッサシステムにおいて、前記複数のノードのそれぞれは、前記メモリにおけるノード間通信用メモリ空間のアドレスが予め固定されており、前記メモリへのアクセス時に前記メモリにおけるアドレスを特定し、該アドレスに基づいて、前記メモリへのアクセスが前記ノード間通信用メモリ空間へのアクセスであるかを判別するメモリアクセス制御部を有することを特徴とするマルチプロセッサシステム。

【請求項8】 請求項7に記載のマルチプロセッサシステムにおいて、

前記メモリアクセス制御部は、前記メモリへのアクセス時に発生した障害を検出するとともに、前記メモリにおける障害が発生した場所のアドレスを特定し、該アドレスに基づいて、前記メモリへのアクセス時に発生した障害が前記ノード間通信用メモリ空間の障害であるかを判別することを特徴とするマルチプロセッサシステム。

【請求項9】 請求項8に記載のマルチプロセッサシステムにおいて、

前記メモリアクセス制御部は、前記ノード間通信用メモ

リ空間において障害が発生したと判断した場合に、前記サービスプロセッサが前記複数のノードのそれぞれに対して障害時の処理内容を通知するための割り込み信号を生成して前記サービスプロセッサに対して出力し、前記サービスプロセッサは、前記割り込み信号が入力されると、該割り込み信号により障害の情報を採取し、該障害の情報に基づく処理内容を、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して通知することを特徴とするマルチプロセッサシステム。

【請求項10】 請求項9に記載のマルチプロセッサシステムにおいて、

前記サービスプロセッサは、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して、ブロードキャスト割り込みを発生して前記処理内容を通知することを特徴とするマルチプロセッサシステム。

【請求項11】 請求項7乃至10のいずれか1項に記載のマルチプロセッサシステムにおいて、

前記複数のノードのそれぞれは、前記セルを複数個有し、該複数のセルのうち、いずれかのセルを分離させることを特徴とするマルチプロセッサシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数のプロセッサ及びメモリからなる複数のノードを有するマルチプロセッサシステムに関し、特に、複数のノードのそれぞれが、内部に設けられたメモリを用いて互いにノード間通信を行うマルチプロセッサシステムに関する。

【0002】

【従来の技術】近年、CPU等のプロセッサが複数設けられたマルチプロセッサシステムにおいては、所定の単位でシステムが分割され、分割されたシステム（以下、ノードと称する）のそれぞれが、独立したマルチプロセッサシステムとして運用可能なように構成されている。

【0003】このようなマルチプロセッサシステムにおいては、ノードのそれぞれが、システムバスに複数のプロセッサ及びメモリが接続されており、複数のプロセッサのそれぞれがメモリを共有することにより、ノードのそれぞれが独立したクラスタシステムとしても運用可能である。

【0004】更に、ノードのそれぞれの内部に設けられたメモリにおいては、その一部が他のノードとの間の通信用として使用されるノード間通信用メモリ空間として定義されており、このノード間通信用メモリ空間を用いて、マルチプロセッサシステムにおける共通のプラットフォーム内で、他のノードとの間で互いに通信が行われる。このような機能は、筐体内クラスタ機能と称される。

【0005】なお、ノードのそれぞれの内部に設けられたメモリにおいては、ノード間通信用メモリ空間と定義された以外の部分が、同じノード内に設けられたプロセ

ッサが共有するためのノード固有メモリ空間と定義される。

【0006】筐体内クラスタ機能を有する一般的なマルチプロセッサシステムにおいては、ノード間通信がプロトコルを用いて行われており、論理プロトコル層にはVIA (Virtual Interconnect Architecture) 等の通信ドライバが用いられ、物理リンク層にEthernetやFC (ファイバーチャネル) 等が用いられる。

【0007】なお、メモリにおけるノード間通信用メモリ空間は、上述した物理リンク層に相当するものである。

【0008】

【発明が解決しようとする課題】従来、ノード間通信用メモリ空間にアクセスする場合、ノード固有メモリ空間にアクセスする場合と同様に、メモリの内部に設けられたアドレスデコード回路を用いてアドレスがそのままデコードされていた。

【0009】しかしながら、アドレスをそのままデコードする場合、アドレスのデコードに多大な時間が消費されたり、アドレスデコード回路の構成が複雑なものとなり、ゲート量の増加やハードウェアのコストアップを招いてしまうという問題点があった。

【0010】また、メモリへのアクセス時に障害が発生した場合、ノード間通信用メモリ空間で発生した障害とノード固有メモリ空間で発生した障害とでプロセッサの障害処理が異なるため、障害が発生したメモリ空間を特定する必要があるが、プロセッサにおいては、ノード間通信用メモリ空間で発生した障害とノード固有メモリ空間で発生した障害との見え方が同じであるため、障害が発生したメモリ空間のアドレスを直接判断する必要があり、これにより、障害処理が複雑なものになってしまうという問題点があった。

【0011】また、メモリへのアクセス時に障害が発生した場合、正常に動作しているノードに対して、障害が発生したメモリが設けられたノードへのアクセスの抑止や障害が発生したメモリが設けられたノードの切り離し等を通知する必要があるが、障害が発生したメモリ空間の判別が遅れると、正常なノードに障害が伝播される可能性が高くなり、システムダウン等が発生してしまうという問題点があった。

【0012】このため、メモリへのアクセス時に障害が発生した場合に、この障害がノード固有メモリ空間における障害及びノード間通信用メモリ空間における障害のいずれであるかを容易に判別することができるマルチプロセッサシステムが望まれていた。

【0013】本発明は、上述したような従来の技術が有する問題点に鑑みてなされたものであって、メモリへのアクセスが、ノード固有メモリ空間へのアクセス及びノード間通信用メモリ空間へのアクセスのいずれであるかを容易に判別することができるマルチプロセッサシステ

ムを提供することを目的とする。

【0014】また、他の目的は、メモリへのアクセス時に障害が発生した場合に、この障害がノード固有メモリ空間における障害及びノード間通信用メモリ空間における障害のいずれであるかを容易に、かつ迅速に判別することができるマルチプロセッサシステムを提供することにある。

【0015】

【課題を解決するための手段】上記目的を達成するために本発明は、複数のプロセッサと、該複数のプロセッサのそれぞれによってデータが記憶されるメモリとを具備するセルからなる複数のノードと、該複数のノードのいずれかにて前記メモリに障害が発生した場合に該複数のノードのそれぞれに対して障害時に処理内容を通知するサービスプロセッサとを有し、前記複数のノードのそれぞれが、前記メモリにおけるメモリ空間の一部をノード間通信用メモリ空間と設定し、前記メモリにおける前記ノード間通信用メモリ空間にアクセスして互いにノード間通信を行うマルチプロセッサシステムにおいて、前記複数のノードのそれぞれは、前記ノード間通信用メモリ空間へのアクセス時に、通信空間アクセスフラグをセットしたトランザクションを発行する通信ドライバと、前記メモリへのアクセス時に前記通信空間アクセスフラグを検出した場合に、前記メモリへのアクセスが前記ノード間通信用メモリ空間へのアクセスであると判断するメモリアクセス制御部とを有することを特徴とする。

【0016】また、前記メモリアクセス制御部は、前記通信ドライバにて発行されたトランザクションを保持するとともに、前記メモリへのアクセス時に発生した障害を検出し、保持された前記トランザクションに前記通信空間アクセスフラグがセットされていることを検出した場合に、前記メモリへのアクセス時に発生した障害が前記ノード間通信用メモリ空間における障害であると判断することを特徴とする。

【0017】また、前記メモリアクセス制御部は、前記ノード間通信用メモリ空間において障害が発生したと判断した場合に、前記サービスプロセッサが前記複数のノードのそれぞれに対して障害時の処理内容を通知するための割り込み信号を生成して前記サービスプロセッサに対して出力し、前記サービスプロセッサは、前記割り込み信号が入力されると、該割り込み信号により障害の情報を採取し、該障害の情報に基づく処理内容を、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して通知することを特徴とする。

【0018】また、前記サービスプロセッサは、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して、ブロードキャスト割り込みを発生して前記処理内容を通知することを特徴とする。

【0019】また、前記複数のノードのそれぞれは、前記セルを複数個有し、該複数のセルのうち、いずれかの

セルを分離させることを特徴とする。

【0020】また、前記メモリアクセス制御部は、前記通信ドライバにて発行されたトランザクションを保持するトランザクション保持回路と、前記メモリへのアクセス時に発生する障害の検出を行うメモリバリティチェック回路と、前記トランザクション保持回路に保持されたトランザクションに前記通信空間アクセスフラグがセットされているか検出するとともに、前記メモリバリティチェック回路にて障害が検出された場合に、前記メモリに発生した障害が前記ノード間通信用メモリ空間における障害であると判断する障害判断部と、前記障害判断部において前記ノード間通信用メモリ空間に障害が発生したと判断された場合に、前記割り込み信号を生成して前記サービスプロセッサに対して出力するメモリ障害割り込みラッチ回路とを有することを特徴とする。

【0021】また、複数のプロセッサと、該複数のプロセッサのそれぞれによってデータが記憶されるメモリとを具備するセルからなる複数のノードと、該複数のノードのいずれかにて前記メモリに障害が発生した場合に該複数のノードのそれぞれに対して障害時に処理内容を通知するサービスプロセッサとを有し、前記複数のノードのそれぞれが、前記メモリにおけるメモリ空間の一部をノード間通信用メモリ空間と設定し、該ノード間通信用メモリ空間にアクセスして互いにノード間通信を行うマルチプロセッサシステムにおいて、前記複数のノードのそれぞれは、前記メモリにおけるノード間通信用メモリ空間のアドレスが予め固定されており、前記メモリへのアクセス時に前記メモリにおけるアドレスを特定し、該アドレスに基づいて、前記メモリへのアクセスが前記ノード間通信用メモリ空間へのアクセスであるかを判別するメモリアクセス制御部を有することを特徴とする。

【0022】また、前記メモリアクセス制御部は、前記メモリへのアクセス時に発生した障害を検出するとともに、前記メモリにおける障害が発生した場所のアドレスを特定し、該アドレスに基づいて、前記メモリへのアクセス時に発生した障害が前記ノード間通信用メモリ空間の障害であるかを判別することを特徴とする。

【0023】また、前記メモリアクセス制御部は、前記ノード間通信用メモリ空間において障害が発生したと判断した場合に、前記サービスプロセッサが前記複数のノードのそれぞれに対して障害時の処理内容を通知するための割り込み信号を生成して前記サービスプロセッサに対して出力し、前記サービスプロセッサは、前記割り込み信号が入力されると、該割り込み信号により障害の情報を採取し、該障害の情報に基づく処理内容を、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して通知することを特徴とする。

【0024】また、前記サービスプロセッサは、前記複数のノードのそれぞれの内部に設けられた複数のプロセッサに対して、ブロードキャスト割り込みを発生して前

記処理内容を通知することを特徴とする。

【0025】また、前記複数のノードのそれぞれは、前記セルを複数個有し、該複数のセルのうち、いずれかのセルを分離させることを特徴とする。

【0026】(作用)上記のように構成された本発明においては、ノードの内部に設けられたメモリにおけるノード間通信用メモリ空間へのアクセス時に、通信ドライバにおいて、通信空間アクセスフラグがセットされたトランザクションが発行され、メモリアクセス制御部において、通信空間アクセスフラグが検出された場合に、ノード間通信用メモリ空間へのアクセスであると判断される。

【0027】また、メモリアクセス制御部においては、通信ドライバにて発行されたトランザクションが保持されるとともに、メモリへのアクセス時に発生した障害が検出され、保持されたトランザクションに通信空間アクセスフラグが検出された場合に、メモリへのアクセス時に発生した障害がノード間通信用メモリ空間の障害であると判断される。

【0028】

【発明の実施の形態】以下に、本発明の実施の形態について図面を参照して説明する。

【0029】図1は、本発明のマルチプロセッサシステムの実施の一形態を示す図である。図1に示すように本形態においては、複数のセル11-1～11-nのそれぞれからなるノード(Node)1-1～1-nと、ノード1-1～1-nのそれぞれの間で行われる通信用の通信路を設定制御するクロスバースイッチ12と、ノード1-1～1-nのいずれかにて障害が発生した場合にノード1-1～1-nのそれぞれに対して処理内容を通知するサービスプロセッサ(診断プロセッサ)13とから構成されており、ノード1-1～1-nのそれぞれが、内部で通信を行うクラスタ機能を有するとともに、ノード1-1～1-nのそれぞれの間で互いに通信を行う筐体内クラスタ機能を有している。

【0030】ノード1-1においては、プロセッサ(Processor)111-1、112-1と、ノード間で行われる通信用のメモリ空間(以下、ノード間通信用メモリ空間と称する)及びノード内で行われる通信用のメモリ空間(以下、ノード固有メモリ空間と称する)を具備し、これらのメモリ空間にてデータのリード及びライトが行われるメモリ(M)113-1と、メモリ113-1におけるノード間通信用メモリ空間へのアクセス時に通信空間アクセスフラグがセットされたトランザクションを発行する通信ドライバ(D)115-1と、通信ドライバ115-1にて発行された通信空間アクセスフラグを含むトランザクションを保持し、メモリ113-1へのアクセスが行われる場合に、この通信空間アクセスフラグに基づいてメモリ113-1へのアクセスがノード間通信用メモリ空間へのアクセスであるかを判別すると

もに、メモリ113-1へのアクセス時に障害が発生した場合に、この通信空間アクセスフラグに基づいてメモリ113-1へのアクセス時に発生した障害がノード間通信用メモリ空間における障害であるかを判別するメモリアクセス制御部であるメモリコントローラ116-1と、データの入出力制御を行うI/O(Input Output)コントローラ114-1とがシステムバス117-1を介して接続されており、メモリコントローラ116-1においては、メモリ113-1へのアクセス時に発生した障害がノード間通信用メモリ空間における障害であると判断された場合、サービスプロセッサ13における障害時の処理内容を決定するための割り込み信号が生成されてサービスプロセッサ13に対して出力される。

【0031】また、ノード1-2～1-nのそれぞれにおいては、プロセッサ111-2～111-n、プロセッサ112-2～112-nと、メモリ113-2～113-nと、メモリコントローラ116-2～116-nと、I/Oコントローラ114-2～114-nと、通信ドライバ115-2～115-nとがシステムバス117-2～117-nを介してそれぞれ接続されており、ノード1-2～1-nのそれぞれに設けられた各部は、ノード1-1内に設けられた各部と同一機能を有している。

【0032】ノード1-1～1-nのそれぞれにおいては、メモリ113-1～113-nにおけるノード間通信用メモリ空間を用いて、通信ドライバ115-1～115-nのそれぞれを介してノード間で互いに通信が行われ、ノード固有メモリ空間を用いて、内部における通信が行われる。

【0033】サービスプロセッサ13は、ノード1-1～1-nのそれぞれの内部に設けられたプロセッサ111-1～111-n、112-1～112-n及びメモリコントローラ116-1～116-nと接続されている。サービスプロセッサ13とプロセッサ111-1～111-n、112-1～112-n及びメモリコントローラ116-1～116-nとの接続方法としては、I/Oコントローラ114-1～114-nまたは通信ドライバ115-1～115-nを介して互いに接続される方法や、互いに直接接続される方法があるが、本形態においては、どちらの接続方法で構成してもよい。

【0034】なお、サービスプロセッサ13は、通常、システム全体のマネージメントを行うものであり、その機能の一部として、ノード1-1～1-nのそれぞれに対する障害時の処理内容の通知を行っている。

【0035】ノード1-1～1-nのそれぞれにおいては、説明を容易にするために2つのプロセッサが設けられた構成にしているが、実際には、更に多くのプロセッサが設けられたノードも存在しているものとする。

【0036】通信ドライバ115-1～115-nのそれぞれにおいては、メモリ113-1～113-nの全

てにおけるノード間通信用メモリ空間のアドレスが認識されている。このようなアドレスを通信ドライバ115-1~115-nのそれぞれに認識させる方法としては、サービスプロセッサ13において、ノード1-1~1-nのそれぞれの組み合わせを、システムを立ち上げる前に通信ドライバ115-1~115-nのそれぞれに設定しておく方法が挙げられる。

【0037】メモリコントローラ116-1~116-nのそれぞれにおいては、メモリ113-1~113-nのそれぞれにおけるメモリ空間のアドレスをデコードするアドレスデコード回路（不図示）と、メモリ113-1~113-nのそれぞれにて障害が発生した場合に、サービスプロセッサ13にて障害処理を行わせるための割り込み信号を生成して出力する割り込み信号生成回路（不図示）とが設けられている。

【0038】本形態においては、ノード1-1~1-nのそれぞれが1つのセルで構成されているが、例えば、ノード1-1~1-nのそれぞれを複数のセルで構成し、セルの一部を筐体内クラスタから分離させることもできる。これにより、ノード1-1~1-nのそれぞれを独立したクラスタシステムとして運用させることが可能になる。

【0039】図2は、図1に示したノード1-1~1-nのそれぞれの内部に設けられたメモリ113-1~113-nにおけるメモリ空間割り当てイメージを示す図である。

【0040】図2においては、メモリ113-1~113-nのそれぞれにおけるノード固有メモリ空間が21-1~21-nであり、メモリ113-1~113-nのそれぞれにおけるノード通信用メモリ空間が22-1~22-nである。

【0041】また、図2における論理アドレス空間イメージは、ノード1-1~1-nの全ノードにおけるメモリ全体のメモリ空間割り当てイメージを示している。

【0042】図2に示すように、メモリ113-1~113-nのそれぞれにおけるノード間通信用メモリ空間は、ノード1-1~1-nのそれぞれがノード固有メモリ空間として使用しない領域に割り当てられる。

【0043】以下に、上記のように構成されたマルチプロセッサシステムにおけるノード間通信動作について説明する。なお、ここでは、ノード1-1、1-2の間で互いにノード間通信を行うものとして、ノード1-1内に設けられたプロセッサ111-1によって、ノード1-2内に設けられたメモリ113-2へのアクセスが行われる場合のノード間通信動作について説明する。

【0044】ノード1-1内に設けられたプロセッサ111-1から、ノード1-2内に設けられたメモリ113-2に対するアクセスが行われると、クロスバスイッチ12が切り替わり、ノード1-1内に設けられた通信ドライバ115-1とノード1-2内に設けられた通

信ドライバ115-2とが接続されて通信路が設定される。

【0045】通信ドライバ115-1においては、メモリ113-1~113-nの全てのメモリにおけるノード間通信用メモリ空間のアドレスが認識されているため、ノード間通信を行う通信路が設定されると、メモリ113-2におけるノード間通信用メモリ空間のアドレスが指定され、指定されたアドレスが通信ドライバ115-2に対して転送される。

【0046】メモリ113-2におけるノード間通信用メモリ空間のアドレスが通信ドライバ115-1から通信ドライバ115-2に転送されると、通信ドライバ115-2において、通信空間アクセスフラグがセットされたトランザクションが発行される。

【0047】図3は、通信空間アクセスフラグがセットされたトランザクションのイメージの一例を示す図である。

【0048】図3に示すように、ノードのそれぞれの間で通信が行われる場合、通信ドライバ115-2において、ビットの一部に通信空間アクセスフラグがセットされたトランザクションが発行される。なお、通信ドライバ115-2にて発行されるトランザクションには、通信ドライバ115-1から転送されたアドレス情報が含まれる。

【0049】次に、メモリコントローラ116-2の内部に設けられた割り込み信号生成回路においては、通信ドライバ115-2にて発行されたトランザクションが保持され、保持されたトランザクションに通信空間アクセスフラグがセットされているかどうか確認される。

【0050】ここで、メモリコントローラ116-2の内部に設けられた割り込み信号生成回路においては、トランザクションに通信空間アクセスフラグがセットされていると確認された場合、メモリ113-2に対するアクセスがノード間通信用メモリ空間へのアクセスであると判断される。

【0051】次に、メモリコントローラ116-2の内部に設けられたアドレスデコード回路において、通信ドライバ115-1から通信ドライバ115-2に転送されたメモリ113-2におけるノード間通信用メモリ空間のアドレスがデコードされる。

【0052】その後、メモリ113-2において、メモリコントローラ116-2内に設けられたアドレスデコード回路にてデコードされたアドレスに、ノード1-1内に設けられたプロセッサ111-1によってデータのリード及びライトが行われる。

【0053】なお、ノード1-1~1-nのうち、どのノードがノード間通信を行う場合も、上記同様のノード間通信動作が行われるものとする。

【0054】上述したように、通信ドライバにおいて、メモリにおけるノード間通信用メモリ空間へのアク

セス時に通信空間アクセスフラグがセットされたトランザクションが発行され、メモリコントローラにおいては、通信ドライバにて発行されたトランザクションに通信空間アクセスフラグがセットされているか確認される。

【0055】このため、メモリへのアクセスが行われる場合、メモリコントローラにおいては、トランザクションに通信空間アクセスフラグがセットされているかを確認するのみで、メモリへのアクセスが、ノード固有メモリ空間へのアクセス及びノード間通信用メモリ空間への

10 アクセスのいずれであるかが判別されることになる。
【0056】また、メモリコントローラにおいては、通信ドライバにて発行されたトランザクションが保持されており、これにより、メモリへのアクセス時に障害が発生した場合においても、メモリコントローラにて保持されたトランザクションに通信空間アクセスフラグがセットされているかを確認することのみで、メモリへのアクセス時に発生した障害が、ノード固有メモリ空間における障害及びノード間通信用メモリ空間における障害のいずれであるかが判別される。

【0057】また、メモリコントローラの内部に設けられたアドレスデコード回路においては、ノード間通信用メモリ空間におけるアドレスをデコードする場合に、メモリに対するアクセスがノード間通信用メモリ空間へのアクセスであると判断された後に、アドレスをデコードすれば良いため、ハードウェア量を削減することが可能になるとともに、ゲート量の削減及びアドレスデコード時間の短縮によりトランザクションのルーティング処理を高速化することが可能になる。

【0058】以下に、ノード内に設けられたメモリへの

30 アクセス時に障害が発生した場合の障害処理動作について詳細に説明する。
【0059】図4は、図1に示したノード1-1内に設けられたメモリ113-1へのアクセス時に障害が発生した場合の障害処理動作を説明するための図である。ここでは、ノード1-1内に設けられたメモリ113-1へのアクセス時にパリティエラー等の障害が発生した場合の障害処理動作について説明するが、ノード1-2～1-nのいずれかの内部に設けられたメモリにパリティエラー等の障害が発生した場合にも同様の障害処理動作が行われるものとする。

【0060】なお、図4に示したマルチプロセッサシステムの構成は、図1に示したマルチプロセッサシステムの構成と同一構成である。

【0061】メモリ113-1へのアクセス時にパリティエラー等の障害が発生すると、メモリコントローラ116-1の内部に設けられた割り込み信号生成回路において、この障害が検出される。

【0062】メモリコントローラ116-1の内部に設けられた割り込み信号生成回路においては、メモリ11

3-1へのアクセスがノード間通信用メモリ空間へのアクセスである場合、通信ドライバ115-1にて発行された通信空間アクセスフラグを含むトランザクションが保持されている。

【0063】このため、メモリコントローラ116-1の内部に設けられた割り込み信号生成回路においては、メモリ113-1における障害を検出すると、通信空間アクセスフラグがセットされているか検出され、通信空間アクセスフラグが検出された場合に、メモリ113-1におけるノード間通信用メモリ空間に障害が発生したと判断される。

【0064】メモリコントローラ116-1の内部に設けられた割り込み信号生成回路においては、メモリ113-1におけるノード間通信用メモリ空間に障害が発生したと判断されると、サービスプロセッサ13に障害処理を行わせるための割り込み信号が生成され、生成された割り込み信号がサービスプロセッサ13に対して出力される。

20 【0065】サービスプロセッサ13においては、メモリコントローラ116-1から出力された割り込み信号が入力されると、入力された割り込み信号に基づいて障害情報が採取され、ノード1-1内に設けられたプロセッサ111-1、112-1に対して障害箇所の切り離し等の障害処理が通知される。

【0066】なお、サービスプロセッサ13においては、メモリ113-1にて発生した障害が重障害である場合、ノード1-1以外のノードの内部に設けられたプロセッサに対し、ノード1-1へのアクセスの抑止や、ノード1-1の切り離し等が通知される。

30 【0067】以下に、図1に示したメモリコントローラ116-1～116-nのそれぞれの内部に設けられた割り込み信号生成回路について詳細に説明する。

【0068】図5は、図1に示したメモリコントローラ116-1の内部に設けられた割り込み信号生成回路の一構成例を示す図である。なお、メモリコントローラ116-2～116-nのそれぞれの内部に設けられた割り込み信号生成回路も同様の構成及び動作が行われるものとする。

40 【0069】本構成例は図5に示すように、通信ドライバ115-1にて発行されたトランザクションを保持するトランザクション保持回路51と、パリティエラー等の障害の検出を行うメモリパリティチェック回路52と、トランザクション保持回路51に保持されたトランザクションにセットされた通信空間アクセスフラグ及びメモリパリティチェック回路52における障害の検出結果に基づいてノード間通信用メモリ空間における障害が発生したか判別する障害判断部53と、障害判断部53にてノード間通信用メモリ空間に障害が発生したと判断された場合に割り込み信号を生成してサービスプロセッサ13に対して出力するメモリ障害割り込みラッチ回路

54と、障害判断部53にてノード間通信用メモリ空間に障害が発生したと判断された場合に、プロセッサ11-1、112-1及びサービスプロセッサ13のうち、メモリ113-1におけるノード固有メモリ空間を用いてデータのリード及びライトが行われているプロセッサに対してノード固有メモリ空間におけるデータのリード及びライトが可能である旨を通知するための信号を生成して出力する障害ステータスレジスタ55とから構成されている。

【0070】以下に、上記のように構成された割り込み信号生成回路の動作について説明する。ここでは、メモリ113-1へのアクセス時にパリティエラーが発生したものと、割り込み信号生成回路の動作について説明する。

【0071】トランザクション保持回路51においては、メモリ113-1へのアクセスがノード間通信用メモリ空間へのアクセスである場合、通信ドライバ115-1にて発行された通信空間アクセスフラグを含むトランザクションが保持される。

【0072】メモリ113-1へのアクセス時にパリティエラーが発生すると、メモリパリティチェック回路52において、メモリ113-1におけるパリティエラーが検出され、該検出結果が障害判断部53に通知される。

【0073】メモリパリティチェック回路52からパリティエラーの検出が通知されると、障害判断部53においては、トランザクション保持回路51に保持されたトランザクションに通信空間アクセスフラグがセットされているかが検出される。

【0074】障害判断部53においては、トランザクション保持回路51にて保持されたトランザクションに通信空間アクセスフラグのビットがセットされていることが検出された場合、メモリ113-1へのアクセス時の障害がノード間通信用メモリ空間の障害であると判断され、該判断結果がメモリ障害割り込みラッチ回路54及び障害ステータスレジスタ55に対して通知される。

【0075】メモリ障害割り込みラッチ回路54においては、障害判断部53からメモリ113-1へのアクセス時に発生した障害がノード間通信用メモリ空間の障害であると通知されると、割り込み信号が生成され、生成された割り込み信号がサービスプロセッサ13に対して出力される。

【0076】また、障害ステータスレジスタ55においては、障害判断部53からメモリ113-1へのアクセス時に発生した障害がノード間通信用メモリ空間の障害であると通知されると、プロセッサ111-1、112-1及びサービスプロセッサ13のうち、ノード固有メモリ空間におけるデータのリード及びライトが行われているプロセッサに対してメモリ113-1におけるノード固有メモリ空間を用いてデータのリード及びライトが

可能である旨を通知するための信号が生成されて出力される。

【0077】一方、障害判断部53において通信空間アクセスフラグのビットがセットされていないと確認された場合、プロセッサ111-1、111-2においては、ノード固有メモリ空間における障害として処理が行われ、特に、2ビットエラーであればシステムダウン等の処理が行われる。

【0078】以下に、ノード間通信を行っているノード内に設けられたメモリにおけるノード間通信用メモリ空間へのアクセス時に障害が発生した場合に、サービスプロセッサ13における障害処理としてブロードキャストを行う場合の障害処理動作について説明する。

【0079】図6は、図1に示したノード1-1内に設けられたメモリ113-1におけるノード間通信用メモリ空間へのアクセス時に障害が発生した場合に、サービスプロセッサ13における障害処理としてブロードキャストを行う場合の障害処理動作を説明するための図である。

【0080】ここでは、メモリ113-1におけるノード間通信用メモリ空間へのアクセス時に障害が発生したことにより、メモリコントローラ116-1にて割り込み信号が生成され、メモリコントローラ116-1にて生成された割り込み信号がサービスプロセッサ13に入力されたものとして、その後の障害処理動作について説明する。

【0081】なお、ノード1-2～1-nのいずれかに設けられたメモリにおけるノード間通信用メモリ空間へのアクセス時に障害が発生した場合にも同様の障害処理動作が行われるものとする。

【0082】また、図6に示したマルチプロセッサシステムは、図1に示したマルチプロセッサシステムの構成と同一構成である。

【0083】図6に示すように、サービスプロセッサ13においては、割り込み信号が入力されると、ノード1-1以外の全てのノードに対してブロードキャストが行われ、障害が発生したメモリ113-1におけるノード間通信用メモリ空間への以後のアクセスが迅速に抑止される。

【0084】これにより、メモリ113-1におけるノード間通信用メモリ空間にて発生した障害が、ノード1-1から他のノードへ伝播されることが最小限に抑止される。

【0085】（他の実施の形態）以下に、他の実施の形態について図1を参照して説明する。

【0086】本形態においては、ノード1-1～1-nのそれぞれが、メモリ113-1～113-nにおけるノード間通信用メモリ空間のアドレスを予め固定した状態でノード間で互いに通信が行われる。

【0087】本形態においては、例えば、ノード1-1

に設けられたメモリ113-1へのアクセス時に障害が発生した場合、メモリコントローラ116-1内に設けられた割り込み信号生成回路においては、メモリ113-1における障害が発生した場所のアドレスが特定され、特定されたアドレスに基づいて、メモリ113-1に発生した障害がノード間通信用メモリ空間に発生した障害であるか判別される。

【0088】なお、本形態においては、メモリ113-1に発生した障害がノード間通信用メモリ空間における障害であると判断された後の障害処理動作は、図1〜図6を用いて説明した実施の形態と同様であるため、詳細な説明は割愛する。

【0089】上述したように本形態においては、図1〜図6を用いて説明した実施の形態と比較して、通信ドライバにおいて通信空間アクセスフラグがセットされたトランザクションを発行させることなく、障害を発生したメモリ空間が、ノード固有メモリ空間及びノード間通信用メモリ空間のいずれであるか判別されるという利点を有しているが、一方では、障害処理時の処理時間が長くなってしまおうという欠点を有している。

【0090】

【発明の効果】以上説明したように本発明においては、ノードの内部に設けられたメモリにおけるノード間通信用メモリ空間へのアクセス時に、通信ドライバにおいて、通信空間アクセスフラグがセットされたトランザクションが発行され、メモリアクセス制御部において、通信空間アクセスフラグが検出された場合にノード間通信用メモリ空間へのアクセスであると判断される。

【0091】このため、メモリへのアクセスがノード間通信用メモリ空間へのアクセス及びノード固有メモリ空間へのアクセスのいずれであるかを容易に判別することができ、これにより、アドレスデコード回路のハードウェア量の低減を図ることができるとともに、ゲート量の削減やデコード時間の短縮が可能となり、トランザクションのルーティング処理の高速化を図ることができる。

【0092】また、メモリアクセス制御部においては、通信ドライバにて発行されたトランザクションが保持されるとともに、メモリへのアクセス時に発生した障害が検出され、保持されたトランザクションに通信空間アクセスフラグがセットされていることが検出された場合に、メモリへのアクセス時に発生した障害がノード間通信用メモリ空間の障害であると判断される。

【0093】このため、メモリへのアクセス時に発生した障害がノード固有メモリ空間及びノード間通信用メモ

リ空間のいずれにおいて発生した障害であるかを容易に、かつ迅速に判別することができ、これにより、ソフトウェアによる障害処理のオーバーヘッドを軽減することができる。

【0094】また、サービスプロセッサにおいては、障害発生時の処理内容が全てのノードに対してブロードキャストして通知するように構成されているため、障害が発生したノードへのアクセスの抑止や、障害が発生したノードの切り離し等が迅速に通知され、これにより、正常なノードへの障害の伝播を未然に防止することができる。

【図面の簡単な説明】

【図1】本発明のマルチプロセッサシステムの実施の一形態を示す図である。

【図2】図1に示したメモリのそれぞれのメモリ空間割り当てイメージを示す図である。

【図3】通信空間アクセスフラグがセットされたトランザクションのイメージの一例を示す図である。

【図4】図1に示したメモリにおけるノード間通信用メモリ空間へのアクセス時に障害が発生した場合の障害処理動作を説明するための図である。

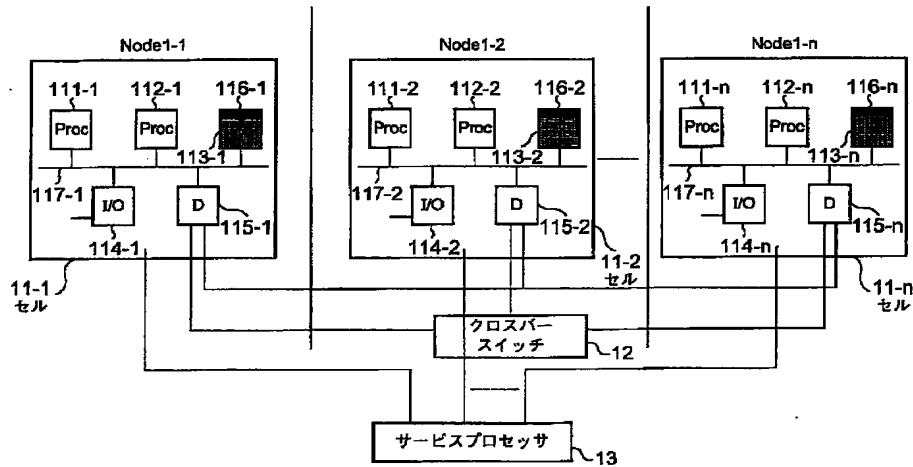
【図5】図1に示したメモリコントローラのそれぞれの内部に設けられた割り込み信号生成回路の一構成例を示す図である。

【図6】図1に示したメモリにおけるノード間通信用メモリ空間へのアクセス時に障害が発生した場合の、他の障害処理動作を説明するための図である。

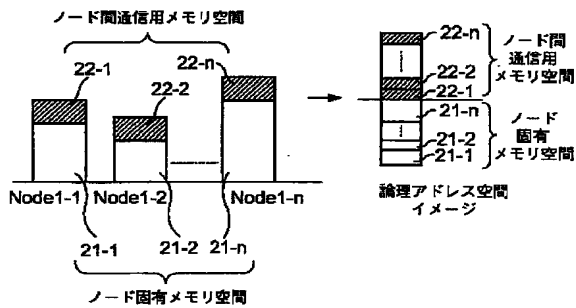
【符号の説明】

- 1-1〜1-n ノード
- 11-1〜11-n セル
- 12 クロスバースイッチ
- 13 サービスプロセッサ
- 111-1〜111-n プロセッサ
- 112-1〜112-n プロセッサ
- 113-1〜113-n メモリ
- 114-1〜114-n I/Oコントローラ
- 115-1〜115-n 通信ドライバ
- 116-1〜116-n メモリコントローラ
- 117-1〜117-n システムバス
- 51 トランザクション保持回路
- 52 メモリバリティチェック回路
- 53 障害判断部
- 54 メモリ障害割り込みラッチ回路
- 55 障害ステータスレジスタ

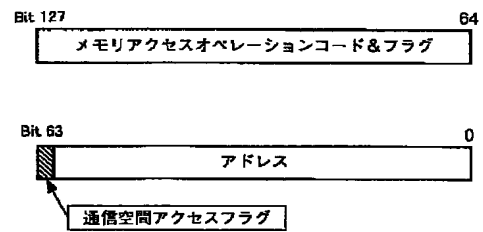
【図1】



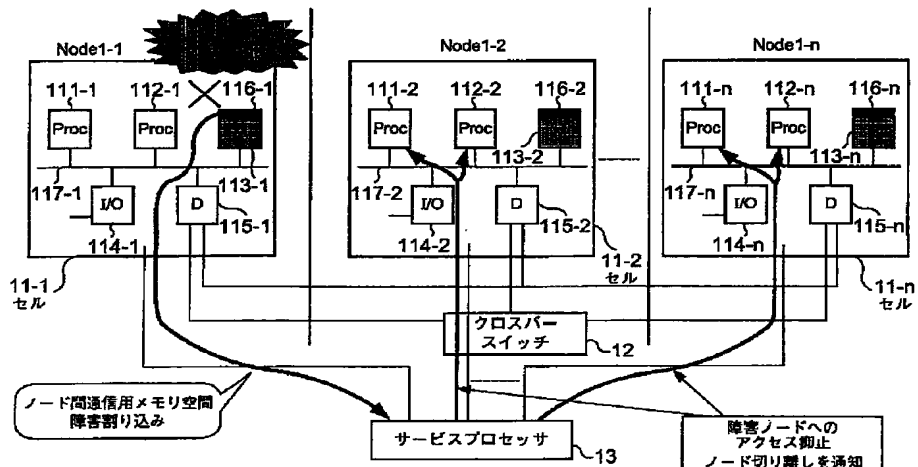
【図2】



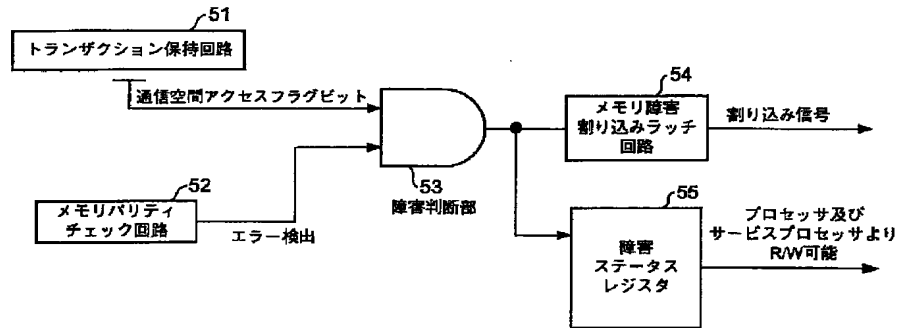
【図3】



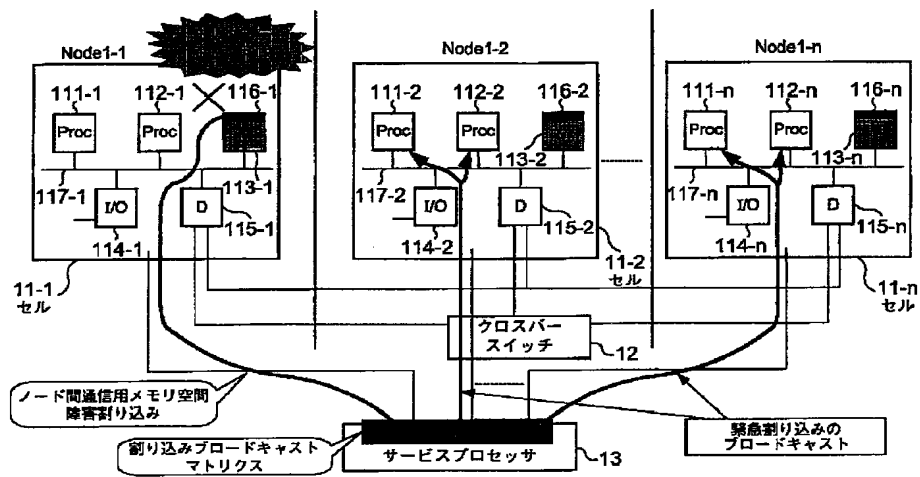
【図4】



【図5】



【図6】



*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] Two or more nodes which consist of a cel possessing the memory by which data are remembered two or more processors and these two or more processors be alike, respectively, It has the service processor which notifies the contents of processing to each of two or more of these nodes at the time of a failure when a failure occurs in said memory in either of these two or more nodes. A part of room [in / respectively / said memory] of two or more of said nodes is set up with the room for the communication link between nodes. In the multiprocessor system which accesses said room for the communication link between nodes in said memory, and performs the communication link between nodes mutually each of two or more of said nodes The communication link driver which publishes the transaction which set the communication link space access flag at the time of access to said room for the communication link between nodes, The multiprocessor system characterized by having the memory access control section judged that access to said memory is access to said room for the communication link between nodes when said communication link space access flag is detected at the time of access to said memory.

[Claim 2] In a multiprocessor system according to claim 1 said memory access control section While holding the transaction published in said communication link driver When it is detected that detect the failure generated at the time of access to said memory, and said communication link space access flag is set to said held transaction The multiprocessor system characterized by judging that the failure generated at the time of access to said memory is a failure in said room for the communication link between nodes.

[Claim 3] In a multiprocessor system according to claim 2 said memory access control

section When it is judged that the failure occurred in said room for the communication link between nodes Generate the interrupt signal for said two or more nodes being alike, respectively, and said service processor receiving, and notifying the contents of processing at the time of a failure, and it outputs to said service processor. Said service processor will extract the information on a failure with this interrupt signal, if said interrupt signal is inputted. The multiprocessor system characterized by notifying the contents of processing based on the information on this failure to two or more processors in which it was prepared inside [each] said two or more nodes.

[Claim 4] It is the multiprocessor system characterized by generating broadcasting interruption and notifying said contents of processing to two or more processors by which said service processor was prepared in each interior of said two or more nodes in the multiprocessor system according to claim 3.

[Claim 5] It is the multiprocessor system characterized by for each of two or more of said nodes having two or more said cels in a multiprocessor system given in claim 1 thru/or any 1 term of 4, and making one of cels separate among these two or more cels.

[Claim 6] In a multiprocessor system given in claim 3 thru/or any 1 term of 5 said memory access control section The transaction holding circuit holding the transaction published in said communication link driver, While detecting whether said communication link space access flag is set to the transaction held in the memory parity check circuit which detects the failure generated at the time of access to said memory, and said transaction holding circuit The failure decision section judged that the failure generated in said memory is a failure in said room for the communication link between nodes when a failure is detected in said memory parity check circuit, The multiprocessor system characterized by having the memory fault interruption latch circuit which generates said interrupt signal and is outputted to said service processor when it is judged that the failure occurred in said room for the communication link between nodes in said failure decision section.

[Claim 7] Two or more nodes which consist of a cel possessing the memory by which data are remembered two or more processors and these two or more processors be alike, respectively, It has the service processor which notifies the contents of processing to each of two or more of these nodes at the time of a failure when a failure occurs in said memory in either of these two or more nodes. In the multiprocessor system which sets up a part of room [in / respectively / said memory] of two or more of said nodes with the room for the communication link between nodes, accesses this room for the communication link between nodes, and

performs the communication link between nodes mutually The address of the room [in / in each of two or more of said nodes / said memory] for the communication link between nodes is being fixed beforehand. The multiprocessor system characterized by having the memory access control section which specifies the address in said memory at the time of access to said memory, and distinguishes whether access to said memory is access to said room for the communication link between nodes based on this address.

[Claim 8] It is the multiprocessor system characterized by distinguishing whether the failure which specified the address of the location which the failure in said memory generated, and was generated based on this address at the time of access to said memory is a failure of said room for the communication link between nodes while detecting the failure which generated said memory access control section in the multiprocessor system according to claim 7 at the time of access to said memory.

[Claim 9] In a multiprocessor system according to claim 8 said memory access control section When it is judged that the failure occurred in said room for the communication link between nodes Generate the interrupt signal for said two or more nodes being alike, respectively, and said service processor receiving, and notifying the contents of processing at the time of a failure, and it outputs to said service processor. Said service processor will extract the information on a failure with this interrupt signal, if said interrupt signal is inputted. The multiprocessor system characterized by notifying the contents of processing based on the information on this failure to two or more processors in which it was prepared inside [each] said two or more nodes.

[Claim 10] It is the multiprocessor system characterized by generating broadcasting interruption and notifying said contents of processing to two or more processors by which said service processor was prepared in each interior of said two or more nodes in the multiprocessor system according to claim 9.

[Claim 11] It is the multiprocessor system characterized by for each of two or more of said nodes having two or more said cels in a multiprocessor system given in claim 7 thru/or any 1 term of 10, and making one of cels separate among these two or more cels.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the multiprocessor system which performs the communication link between nodes mutually especially using the memory of two or more nodes prepared in the interior, respectively about the multiprocessor system which has two or more nodes which consist of two or more processors and memory.

[0002]

[Description of the Prior Art] It is constituted so that it can apply in recent years as a multiprocessor system with which the system (a node is called hereafter) which the system was divided and was divided in the predetermined unit in the multiprocessor system with which two or more processors, such as CPU, were prepared became independent, respectively.

[0003] In such a multiprocessor system, the processor of plurality [system bus] respectively of a node and memory are connected, and when each of two or more processors shares memory, it can apply also as a cluster system with which each of a node became independent.

[0004] Furthermore, in the memory prepared in each interior of a node, it defines as room for the communication link between nodes where that part is used as an object for the communication link between other nodes, and a communication link is performed mutually among other nodes in the common platform in a multiprocessor system using this room for the communication link between nodes. Such a function is called the cluster ability in a case.

[0005] In addition, the memory prepared in each interior of a node is defined as node proper room for the processor by which the part except having been defined as the room for the communication link between nodes was prepared in the same node to

share.

[0006] In the general multiprocessor system which has the cluster ability in a case, the communication link between nodes is performed using the protocol, communication link drivers, such as VIA (Virtual Interconnect Architecture), are used for a logical protocol layer, and Ether, FC (Fibre Channel), etc. are used for a physical-link layer.

[0007] In addition, the room for the communication link between nodes in memory is equivalent to the physical-link layer mentioned above.

[0008]

[Problem(s) to be Solved by the Invention] When the room for the communication link between nodes was accessed conventionally, the address was decoded as it was like the case where node proper room is accessed, using the address decoding circuit established in the interior of memory.

[0009] However, when the address was decoded as it was, great time amount was consumed by decoding of the address, or the configuration of an address decoding circuit became complicated, and there was a trouble of causing the increment in the amount of gates and the cost rise of hardware.

[0010] Moreover, although it is necessary to pinpoint the room which the failure generated since error processing of a processor differs with the failure generated in the room for the communication link between nodes, and the failure generated in node proper room when a failure occurs at the time of access to memory Since how of the failure generated in the room for the communication link between nodes and the failure generated in node proper room to be visible is the same in a processor, The address of the room which the failure generated needed to be judged directly and there was a trouble that error processing will become complicated by this.

[0011] Moreover, although it is necessary to notify separation of the node in which the memory which suppression and failure of access to the node in which the memory which the failure generated was prepared generated was prepared etc. to the node which is operating normally when a failure occurs at the time of access to memory When distinction of the room which the failure generated was overdue, possibility that a failure spread to a normal node became high, and there was a trouble that a system down etc. will occur.

[0012] For this reason, when a failure occurred at the time of access to memory, a multiprocessor system which can be distinguished easily was desired [any of the failure in node proper room, and the failure in the room for the communication link between nodes this failure is, and].

[0013] This invention is made in view of the trouble which a Prior art which was mentioned above has, and aims at offering the multiprocessor system which can distinguish easily any of access to node proper room, and access to the room for the communication link between nodes access to memory is.

[0014] Moreover, other purposes have this failure in offering the multiprocessor system which can distinguish easily and quickly any of the failure in node proper room, and the failure in the room for the communication link between nodes they are, when a failure occurs at the time of access to memory.

[0015]

[Means for Solving the Problem] Two or more nodes which consist of a cel possessing the memory by which data are remembered that two or more processors and these two or more processors should boil this invention, respectively in order to attain the above-mentioned purpose, It has the service processor which notifies the contents of processing to each of two or more of these nodes at the time of a failure when a failure occurs in said memory in either of these two or more nodes. A part of room [in / respectively / said memory] of two or more of said nodes is set up with the room for the communication link between nodes. In the multiprocessor system which accesses said room for the communication link between nodes in said memory, and performs the communication link between nodes mutually each of two or more of said nodes The communication link driver which publishes the transaction which set the communication link space access flag at the time of access to said room for the communication link between nodes, When said communication link space access flag is detected at the time of access to said memory, it is characterized by having the memory access control section judged that access to said memory is access to said room for the communication link between nodes.

[0016] Moreover, while said memory-access control section holds the transaction published in said communication link driver, when it is detected that detect the failure generated at the time of access to said memory, and said communication link space access flag is set to said held transaction, it is characterized by to judge that the failure generated at the time of access to said memory is a failure in said room for the communication link between nodes.

[0017] Moreover, when it is judged that the failure generated said memory access control section in said room for the communication link between nodes Generate the interrupt signal for said two or more nodes being alike, respectively, and said service processor receiving, and notifying the contents of processing at the time of a failure, and it outputs to said service processor. If said interrupt signal is inputted, said

service processor will extract the information on a failure with this interrupt signal, and will be characterized by notifying the contents of processing based on the information on this failure to two or more processors in which it was prepared inside [each] said two or more nodes.

[0018] Moreover, said service processor is characterized by generating broadcasting interruption and notifying said contents of processing to two or more processors prepared in each interior of said two or more nodes.

[0019] Moreover, each of two or more of said nodes has two or more said cels, and is characterized by making one of cels separate among these two or more cels.

[0020] Moreover, the transaction holding circuit holding the transaction from which said memory access control section was published in said communication link driver, While detecting whether said communication link space access flag is set to the transaction held in the memory parity check circuit which detects the failure generated at the time of access to said memory, and said transaction holding circuit The failure decision section judged that the failure generated in said memory is a failure in said room for the communication link between nodes when a failure is detected in said memory parity check circuit, When it is judged that the failure occurred in said room for the communication link between nodes in said failure decision section, it is characterized by having the memory fault interruption latch circuit which generates said interrupt signal and is outputted to said service processor.

[0021] Moreover, two or more nodes which consist of a cel possessing the memory by which data are remembered two or more processors and these two or more processors be alike, respectively, It has the service processor which notifies the contents of processing to each of two or more of these nodes at the time of a failure when a failure occurs in said memory in either of these two or more nodes. In the multiprocessor system which sets up a part of room [in / respectively / said memory] of two or more of said nodes with the room for the communication link between nodes, accesses this room for the communication link between nodes, and performs the communication link between nodes mutually The address of the room [in / in each of two or more of said nodes / said memory] for the communication link between nodes is being fixed beforehand. The address in said memory is specified at the time of access to said memory, and it is characterized by having the memory access control section which distinguishes whether access to said memory is access to said room for the communication link between nodes based on this address.

[0022] Moreover, said memory access control section specifies the address of the

location which the failure in said memory generated, and is characterized by distinguishing whether the failure generated at the time of access to said memory is a failure of said room for the communication link between nodes based on this address while it detects the failure generated at the time of access to said memory.

[0023] Moreover, when it is judged that the failure generated said memory access control section in said room for the communication link between nodes Generate the interrupt signal for said two or more nodes being alike, respectively, and said service processor receiving, and notifying the contents of processing at the time of a failure, and it outputs to said service processor. If said interrupt signal is inputted, said service processor will extract the information on a failure with this interrupt signal, and will be characterized by notifying the contents of processing based on the information on this failure to two or more processors in which it was prepared inside [each] said two or more nodes.

[0024] Moreover, said service processor is characterized by generating broadcasting interruption and notifying said contents of processing to two or more processors prepared in each interior of said two or more nodes.

[0025] Moreover, each of two or more of said nodes has two or more said cels, and is characterized by making one of cels separate among these two or more cels.

[0026] (Operation) In this invention constituted as mentioned above, the transaction by which the communication link space access flag was set in the communication link driver at the time of access to the room for the communication link between nodes in the memory prepared in the interior of a node is published, and when a communication link space access flag is detected in a memory access control section, it is judged that it is access to the room for the communication link between nodes.

[0027] Moreover, while the transaction published in the communication link driver is held in a memory access control section, when the failure generated at the time of access to memory is detected and a communication link space access flag is detected by the held transaction, the failure generated at the time of access to memory is judged to be the failure of the room for the communication link between nodes.

[0028]

[Embodiment of the Invention] Below, the gestalt of operation of this invention is explained with reference to a drawing.

[0029] Drawing 1 is drawing showing one gestalt of operation of the multiprocessor system of this invention. The node (Node) 1-1 which consists of each of two or more cels 11-1 - 11-n in this gestalt as shown in drawing 1 - 1-n, The crossbar switch 12 which carries out setting control of the channel for a communication link performed

between each of a node 1-1 - 1-n, It consists of service processors (diagnostic processor) 13 which notify the contents of processing to each of a node 1-1 - 1-n when a failure occurs in either a node 1-1 - 1-n. While having the cluster ability of a node 1-1 - 1-n which communicates inside, respectively, it has the cluster ability in a case which communicates mutually between each of a node 1-1 - 1-n.

[0030] In a node 1-1 A processor 111-1,112-1 (Proc), The room for a communication link performed within the room for a communication link (the room for the communication link between nodes is called hereafter) performed between nodes, and a node (Memory M) 113-1 to which it provides (node proper room is called hereafter), and a lead and light of data are performed in such rooms, Communication link (driver D) 115-1 which publishes the transaction by which the communication link space access flag was set at the time of access to the room for the communication link between nodes in memory 113-1, The transaction containing the communication link space access flag published by the communication link driver 115-1 is held. When access to memory 113-1 is performed, while distinguishing whether access to memory 113-1 is access to the room for the communication link between nodes based on this communication link space access flag When a failure occurs at the time of access to memory 113-1 The memory controller 116-1 which is the memory access control section which distinguishes whether the failure generated based on this communication link space access flag at the time of access to memory 113-1 is a failure in the room for the communication link between nodes, The I/O (Input Output) controller 114-1 which performs input/output control of data is connected through the system bus 117-1. When the failure generated at the time of access to memory 113-1 is judged to be a failure in the room for the communication link between nodes in the memory controller 116-1, The interrupt signal for determining the contents of processing at the time of the failure in a service processor 13 is generated, and it is outputted to a service processor 13.

[0031] Moreover, it sets to each of a node 1-2 - 1-n. A processor 111-2 - 111-n, and a processor 112-2 - 112-n, Memory 113-2 - 113-n, and the memory controller 116-2 - 116-n, I/O controller 114-2 - 114-n, and the communication link driver 115-2 - 115-n are connected through a system bus 117-2 - 117-n, respectively. Each part prepared in each of a node 1-2 - 1-n has the same function as each part prepared in the node 1-1.

[0032] In each of a node 1-1 - 1-n, using the room for the communication link between nodes in memory 113-1 - 113-n, a communication link is mutually performed between nodes through each of the communication link driver 115-1 - 115-n, and the

communication link in the interior is performed using node proper room.

[0033] The service processor 13 is connected with the processor 111-1 prepared in each interior of a node 1-1 - 1-n - 111-n, 112-1 - 112-n, and the memory controller 116-1 - 116-n. Although there are an approach mutually connected through I/O controller 114-1 - 114-n, or the communication link driver 115-1 - 115-n as a connection method with a service processor 13, a processor 111-1 - 111-n, 112-1 - 112-n, and the memory controller 116-1 - 116-n and an approach by which direct continuation is carried out mutually, in this gestalt, you may constitute from a which connection method.

[0034] In addition, a service processor 13 manages the whole system and is usually notifying the contents of processing at the time of the failure over each of a node 1-1 - 1-n as a part of the function.

[0035] In each of a node 1-1 - 1-n, although it is made the configuration in which two processors were prepared in order to give explanation easy, the node in which actual further many processors were prepared shall also exist.

[0036] In each of the communication link driver 115-1 - 115-n, the address of the room for the communication link between nodes in all of memory 113-1 - 113-n is recognized. The method of setting each combination of a node 1-1 - 1-n as each of the communication link driver 115-1 - 115-n in the service processor 13, as an approach of making each of the communication link driver 115-1 - 115-n recognizing such the address, before starting a system is mentioned.

[0037] In each of the memory controller 116-1 - 116-n, when the address decoding circuit (un-illustrating) which decodes the address of the room in each of memory 113-1 - 113-n, and memory 113-1 - 113-n are alike, respectively and a failure occurs, the interrupt signal generation circuit (un-illustrating) which generates and outputs the interrupt signal for making error processing perform by the service processor 13 is prepared.

[0038] Although each of a node 1-1 - 1-n consists of one cel, for example, each of a node 1-1 - 1-n can be constituted from two or more cels, and a part of cel can also be made to separate from the cluster in a case in this gestalt. It becomes possible to make each of a node 1-1 - 1-n by this employ as an independent cluster system.

[0039] Drawing 2 is drawing showing the room quota image in the memory 113-1 prepared in each interior of the node 1-1 shown in drawing 1 - 1-n - 113-n.

[0040] In drawing 2, the node proper rooms in each of memory 113-1 - 113-n are 21-1 - 21-n, and the rooms for a node communication link in each of memory 113-1 - 113-n are 22-1 - 22-n.

[0041] Moreover, the logical address space image in drawing 2 shows the room quota image of the whole memory in all the nodes of a node 1-1 - 1-n.

[0042] As shown in drawing 2, the room for the communication link between nodes in each of memory 113-1 - 113-n is assigned to the field which each of a node 1-1 - 1-n does not use as node proper room.

[0043] Below, the node communication link actuation in the multiprocessor system constituted as mentioned above is explained. In addition, the processor 111-1 prepared in the node 1-1 explains node communication link actuation in case access to the memory 113-2 prepared in the node 1-2 is performed here as a node 1-1 and a thing which performs the communication link between nodes mutually among 1-2.

[0044] If access to the memory 113-2 prepared in the node 1-2 is performed from the processor 111-1 prepared in the node 1-1, a crossbar switch 12 will change, the communication link driver 115-1 prepared in the node 1-1 and the communication link driver 115-2 prepared in the node 1-2 will be connected, and a channel will be set up.

[0045] In the communication link driver 115-1, since the address of the room for the communication link between nodes in all the memory of memory 113-1 - 113-n is recognized, if the channel which performs the communication link between nodes is set up, the address of the room for the communication link between nodes in memory 113-2 will be specified, and the specified address will be transmitted to the communication link driver 115-2.

[0046] If the address of the room for the communication link between nodes in memory 113-2 is transmitted to the communication link driver 115-2 from the communication link driver 115-1, in the communication link driver 115-2, the transaction to which the communication link space access flag was set will be published.

[0047] Drawing 3 is drawing showing an example of the image of the transaction to which the communication link space access flag was set.

[0048] As shown in drawing 3, when a communication link is performed between each of a node, in the communication link driver 115-2, the transaction by which the communication link space access flag was set to a part of bit is published. In addition, the address information transmitted from the communication link driver 115-1 is contained in the transaction published by the communication link driver 115-2.

[0049] Next, in the interrupt signal generation circuit established in the interior of the memory controller 116-2, it is checked whether the transaction published by the communication link driver 115-2 is held, and the communication link space access flag is set to the held transaction.

[0050] Here, in the interrupt signal generation circuit established in the interior of the memory controller 116-2, when the communication link space access flag was set to the transaction and it is checked, access to memory 113-2 is judged to be access to the room for the communication link between nodes.

[0051] Next, in the address decoding circuit established in the interior of the memory controller 116-2, the address of the room for the communication link between nodes in the memory 113-2 transmitted to the communication link driver 115-2 from the communication link driver 115-1 is decoded.

[0052] Then, a lead and light of data are performed in memory 113-2 by the processor 111-1 prepared in the address decoded in the address decoding circuit prepared in the memory controller 116-2 in the node 1-1.

[0053] In addition, when every node performs the communication link between nodes among a node 1-1 - 1-n, the same node communication link actuation as the above shall be performed.

[0054] As mentioned above, in a communication link driver, it is checked whether the communication link space access flag is set to the transaction which the transaction by which the communication link space access flag was set at the time of access to the room for the communication link between nodes in memory was published, and was published in the communication link driver in the memory controller.

[0055] For this reason, when access to memory is performed, in a memory controller, it will be distinguished any of access to node proper room and access to the room for the communication link between nodes it is only checking whether the communication link space access flag being set to the transaction, and access to memory is.

[0056] In the memory controller, the transaction published in the communication link driver is held. Moreover, by this Only by checking whether the communication link space access flag is set to the transaction held by the memory controller, when a failure occurs at the time of access to memory It is distinguished any of the failure in node proper room and the failure in the room for the communication link between nodes the failures generated at the time of access to memory are.

[0057] Moreover, in the address decoding circuit established in the interior of a memory controller, while becoming possible to reduce the amount of hardware in order for what is necessary to be just to decode the address after access to memory is judged to be access to the room for the communication link between nodes when decoding the address in the room for the communication link between nodes, it becomes possible to accelerate the routing processing of a transaction by reduction of the amount of gates, and compaction of address decoding time amount.

[0058] Below, error-processing actuation when a failure occurs at the time of access to the memory prepared in the node is explained at a detail.

[0059] Drawing 4 is drawing for explaining error-processing actuation when a failure occurs at the time of access to the memory 113-1 prepared in the node 1-1 shown in drawing 1 . Here, although error-processing actuation when failures, such as a parity error, occur at the time of access to the memory 113-1 prepared in the node 1-1 is explained, also when failures, such as a parity error, occur in the memory prepared in the interior of either a node 1-2 - 1-n, same error-processing actuation shall be performed.

[0060] In addition, the configuration of the multiprocessor system shown in drawing 4 is the same configuration as the configuration of the multiprocessor system shown in drawing 1 .

[0061] If failures, such as a parity error, occur at the time of access to memory 113-1, this failure will be detected in the interrupt signal generation circuit established in the interior of the memory controller 116-1.

[0062] In the interrupt signal generation circuit established in the interior of the memory controller 116-1, when access to memory 113-1 is access to the room for the communication link between nodes, the transaction containing the communication link space access flag published by the communication link driver 115-1 is held.

[0063] For this reason, in the interrupt signal generation circuit established in the interior of the memory controller 116-1, if the failure in memory 113-1 is detected, it will be judged that the failure occurred in the room for the communication link between nodes in memory 113-1 whether the communication link space access flag is set, when it is detected and a communication link space access flag is detected.

[0064] In the interrupt signal generation circuit established in the interior of the memory controller 116-1, if it is judged that the failure occurred in the room for the communication link between nodes in memory 113-1, the interrupt signal for making error processing perform will be generated by the service processor 13, and the generated interrupt signal will be outputted to it to a service processor 13.

[0065] In a service processor 13, if the interrupt signal outputted from the memory controller 116-1 is inputted, fault information will be extracted based on the inputted interrupt signal, and error processing, such as separation of a failure part, will be notified to the processor 111-1, 112-1 prepared in the node 1-1.

[0066] In addition, in a service processor 13, when the failure generated by memory 113-1 is a heavy failure, suppression of access, separation of the node 1-1 to a node 1-1, etc. are notified to the processor prepared in the interior of nodes other than

node 1-1.

[0067] The interrupt signal generation circuit established in each interior of the memory controller 116-1 shown below at drawing 1 - 116-n is explained to a detail.

[0068] Drawing 5 is drawing showing the example of 1 configuration of the interrupt signal generation circuit established in the interior of the memory controller 116-1 shown in drawing 1. In addition, a configuration and actuation with the same said of the interrupt signal generation circuit established in each interior of the memory controller 116-2 - 116-n shall be performed.

[0069] The transaction holding circuit 51 which holds the transaction published by the communication link driver 115-1 as this example of a configuration is shown in drawing 5, The memory parity check circuit 52 which detects failures, such as a parity error, Whether based on the detection result of the failure in the communication link space access flag and the memory parity check circuit 52 which were set to the transaction held in the transaction holding circuit 51, the failure in the room for the communication link between nodes occurred, and the failure decision section 53 to distinguish, The memory fault interruption latch circuit 54 which generates an interrupt signal and is outputted to a service processor 13 when it is judged that the failure occurred in the room for the communication link between nodes in the failure decision section 53, When it is judged that the failure occurred in the room for the communication link between nodes in the failure decision section 53 The inside of a processor 111-1, 112-1 and a service processor 13, As opposed to the processor to which a lead and light of data are performed using the node proper room in memory 113-1 It consists of failure status registers 55 which generate and output the signal for notifying the purport in which the lead and light of data in node proper room are possible.

[0070] Below, actuation of the interrupt signal generation circuit constituted as mentioned above is explained. Here, actuation of an interrupt signal generation circuit is explained as what the parity error generated at the time of access to memory 113-1.

[0071] In the transaction holding circuit 51, when access to memory 113-1 is access to the room for the communication link between nodes, the transaction containing the communication link space access flag published by the communication link driver 115-1 is held.

[0072] If a parity error occurs at the time of access to memory 113-1, in the memory parity check circuit 52, the parity error in memory 113-1 will be detected, and this detection result will be notified to the failure decision section 53.

[0073] If detection of a parity error is notified from the memory parity check circuit 52,

in the failure decision section 53, it will be detected whether the communication link space access flag is set to the transaction held in the transaction holding circuit 51.

[0074] In the failure decision section 53, when it is detected that the bit of a communication link space access flag is set to the transaction held in the transaction holding circuit 51, the failure at the time of access to memory 113-1 is judged to be the failure of the room for the communication link between nodes, and this decision result is notified to the memory fault interruption latch circuit 54 and the failure status register 55.

[0075] In the memory fault interruption latch circuit 54, if notified that the failure generated at the time of access to memory 113-1 from the failure decision section 53 is a failure of the room for the communication link between nodes, an interrupt signal will be generated and the generated interrupt signal will be outputted to a service processor 13.

[0076] Moreover, it sets to the failure status register 55. If notified that the failure generated at the time of access to memory 113-1 from the failure decision section 53 is a failure of the room for the communication link between nodes The inside of a processor 111-1, 112-1 and a service processor 13, The signal for notifying the purport in which a lead and light of data are possible using the node proper room in memory 113-1 to the processor to which the lead and light of data in node proper room are performed is generated and outputted.

[0077] On the other hand, when the bit of a communication link space access flag was not set in the failure decision section 53 and it is checked, in a processor 111-1, 111-2, processing is performed as a failure in node proper room, and if it is two bit errors, processing of a system down etc. will be performed especially.

[0078] When a failure occurs at the time of access to the room for the communication link between nodes in the memory prepared in the node which is performing the communication link between nodes to below, the error-processing actuation in the case of broadcasting as error processing in a service processor 13 is explained.

[0079] Drawing 6 is drawing for explaining the error-processing actuation in the case of broadcasting as error processing in a service processor 13, when a failure occurs at the time of access to the room for the communication link between nodes in the memory 113-1 prepared in the node 1-1 shown in drawing 1 .

[0080] Here, when the failure occurred at the time of access to the room for the communication link between nodes in memory 113-1 explains subsequent error-processing actuation as that as which the interrupt signal which the interrupt signal was generated by the memory controller 116-1, and was generated by the

memory controller 116-1 was inputted into the service processor 13.

[0081] In addition, also when a failure occurs at the time of access to the room for the communication link between nodes in the memory prepared in either a node 1-2 - 1-n, same error-processing actuation shall be performed.

[0082] Moreover, the multiprocessor system shown in drawing 6 is the same configuration as the configuration of the multiprocessor system shown in drawing 1 .

[0083] As shown in drawing 6 , if an interrupt signal is inputted in a service processor 13, broadcasting will be performed to all nodes other than node 1-1, and access after the room for the communication link between nodes in the memory 113-1 which the failure generated will be inhibited quickly.

[0084] It is inhibited to the minimum that the failure generated in the room for the communication link between nodes in memory 113-1 spreads from a node 1-1 to other nodes by this.

[0085] (Gestalt of other operations) Below, the gestalt of other operations is explained with reference to drawing 1 .

[0086] In this gestalt, where the address of the room [in / respectively / memory 113-1 - 113-n] for the communication link between nodes of a node 1-1 - 1-n is fixed beforehand, a communication link is mutually performed between nodes.

[0087] When a failure occurs in this gestalt at the time of access to the memory 113-1 prepared in the node 1-1, for example, In the interrupt signal generation circuit prepared in the memory controller 116-1 It is distinguished whether the failure which the address of the location which the failure in memory 113-1 generated was specified, and was generated in memory 113-1 based on the specified address is a failure generated in the room for the communication link between nodes.

[0088] In addition, in this gestalt, since the error-processing actuation after the failure generated in memory 113-1 was judged to be a failure in the room for the communication link between nodes is the same as that of the gestalt of the operation explained using drawing 1 - drawing 6 , detailed explanation is omitted.

[0089] As mentioned above, it compares with the gestalt of the operation explained using drawing 1 - drawing 6 in this gestalt. Although it has any of node proper room and the room for the communication link between nodes the room which generated the failure is, and the advantage of being distinguished, without making the transaction to which the communication link space access flag was set in the communication link driver publish On the other hand, it has the fault that the processing time at the time of error processing will become long.

[0090]

[Effect of the Invention] As explained above, the transaction by which the communication link space access flag was set in the communication link driver in this invention at the time of access to the room for the communication link between nodes in the memory prepared in the interior of a node is published, and in a memory access control section, when a communication link space access flag is detected, it is judged that it is access to the room for the communication link between nodes.

[0091] For this reason, it can distinguish easily any of access to the room for the communication link between nodes, and access to node proper room access to memory is, and thereby, while being able to aim at reduction of the amount of hardware of an address decoding circuit, reduction of the amount of gates and compaction of decoding time amount are attained, and improvement in the speed of the routing processing of a transaction can be attained.

[0092] Moreover, while the transaction published in the communication link driver is held in a memory access control section, when it is detected that the failure generated at the time of access to memory is detected, and the communication link space access flag is set to the held transaction, the failure generated at the time of access to memory is judged to be the failure of the room for the communication link between nodes.

[0093] For this reason, the failure generated at the time of access to memory can distinguish easily and quickly whether it is the failure generated in any of node proper room and the room for the communication link between nodes, and, thereby, can mitigate the overhead of error processing by software.

[0094] Moreover, in a service processor, since it is constituted so that the contents of processing at the time of failure generating may broadcast and notify to all nodes, suppression of access to the node which the failure generated, separation of the node which the failure generated, etc. are notified quickly, and, thereby, can prevent propagation of the failure to a normal node beforehand.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is drawing showing one gestalt of operation of the multiprocessor system of this invention.

[Drawing 2] It is drawing showing each room quota image of the memory shown in drawing 1 .

[Drawing 3] It is drawing showing an example of the image of the transaction to which the communication link space access flag was set.

[Drawing 4] It is drawing for explaining error-processing actuation when a failure occurs at the time of access to the room for the communication link between nodes in the memory shown in drawing 1 .

[Drawing 5] It is drawing showing the example of 1 configuration of the interrupt signal generation circuit established in each interior of the memory controller shown in drawing 1 .

[Drawing 6] It is drawing for explaining other error-processing actuation when a failure occurs at the time of access to the room for the communication link between nodes in the memory shown in drawing 1 .

[Description of Notations]

1-1 - 1-n Node

11-1 - 11-n Cel

12 Crossbar Switch

13 Service Processor

111-1 - 111-n Processor

112-1 - 112-n Processor

113-1 - 113-n Memory

114-1 - 114-n I/O controller

115-1 - 115-n Communication link driver

116-1 - 116-n Memory controller

117-1 - 117-n System bus

- 51 Transaction Holding Circuit
- 52 Memory Parity Check Circuit
- 53 Failure Decision Section
- 54 Memory Fault Interruption Latch Circuit
- 55 Failure Status Register

[Translation done.]